INTERNATIONAL UNIVERSITY OF JAPAN
Public Management and Policy Analysis Program
Graduate School of International Relations

ADC6512
Topics in Data Analysis (Panel Data Models Using Stata)
(2 Credits)
Winter 2012

Classroom: C204
Time: 14:40-17:50 on Friday
Web : http://www.sonsoo.org/method/panel/
E-mail: kucc625@iuj.ac.jp

Instructor: Hun Myoung Park
Office: 328
Office Hour: Wed. 15:00-16:00
Telephone: (025) 779-1424

Prerequisites: Students are assumed to have basic understandings of statistics and ordinary least squares. Also college level (matrix) algebra and some calculus are required. The target audience is (2nd year) linkages students in IDP and PMPP who have already taken *Statistics for Economics and Management* (DCC5220) and *Applied Econometrics* (DCC 5260) but have not taken *Cross-sectional and Panel Analysis* (ADC6515).

INTRODUCTION

Panel data are also called *longitudinal data* or *cross-sectional and time-series data*. A panel data set has multiple entities, each of which has repeated measurements at different time periods. Panel data may have *individual effect*, *time effect*, or both, which are analyzed by *fixed effect* and/or *random effect* models.

As more and more panel data are available, many scholars, practitioners, and students have been interested in panel data modeling because these longitudinal data have more variability and allow to explore more issues than do cross-sectional or time-series data alone (Kennedy, 2008: 282). Baltagi (2008) puts, "Panel data give more informative data, more variability, less collinearity among the variables, more degrees of freedom and more efficiency" (p.7). Given well-organized panel data, panel data models are definitely attractive and appealing since they provide ways of dealing with heterogeneity and examine fixed and/or random effects in the longitudinal data.

However, panel data modeling is not as easy as it sounds. A common misunderstanding is that fixed and/or random effect models should always be employed whenever your data are arranged in the panel data format. The problems of panel data modeling come from 1) panel data themselves, 2) modeling process, and 3) interpretation and presentation of the result. Some studies analyze poorly organized panel data (in fact, they are not longitudinal in a strong econometric sense) and others mechanically apply fixed and/or random effect models in haste without consideration of relevance of such models. Careless researchers often fail to interpret the results correctly and to present them appropriately.

This course is designed to provide 1) basic understandings of panel data models, 2) practical (as opposed to theoretical) guides of panel data modeling with special emphasis on substantive interpretation and professional presentation of results, and 3) practical knowledge and skills in manipulating data and running econometric models with Stata (scientific data

analysis and replicability of research will be emphasized). This "gentle introduction to panel data models" does not, however, touch upon such advanced topics as nonlinear panel data model (logit/probit models), dynamic panel data models, and model for parameter heterogeneity (hierarchical linear model), which should be addressed in a doctoral level course. This class will be ideal, in particular, for linkage students of IDP and PMPP who want to use panel data models for their master's theses.

At the end of this term, students should be able to:

a. Understand problems of OLS when analyzing panel data,
b. Understand basics of panel data models,
c. Understand hypothesis tests in panel data models and their limits,
d. Organize, clean, and manipulate panel data using Stata,
e. Analyze properties of panel data,
f. Understand pros and cons of three approaches to LSDV,
g. Understand pros and cons of "Within" estimation,
h. Understand natures of fixed and random effect models,
i. Choose a proper panel data model for their data,
j. Read and report Stata output correctly,
k. Interpret the result substantively,
l. Present the result in a professional manner, and
m. Use Stata for econometric data analysis

Related courses are *Cross-sectional and Panel Analysis* (ADC6515) and *Time Series Analysis* (ADC 6525). The former targets regular IDP students with emphasis on theory, while this course focuses more on practical applications of panel data models.

ORGANIZATION

This course is built on lectures, labs, homework assignments, and term project, among which the last two are most important. Active participation in class discussion is very important to both instructor and students. It is highly recommended that students take advantage of talking to the instructor during office hours or by setting up appointments.

ASSESSMENT (COURSE REQUIREMENT)

Attendance and Participation: Students should attend ALL classes and actively participate in class, labs, and group project. Each unexcused absence results in THREE POINT deduction from the final score. Extra credits may be given to students who make an outstanding contribution to class and who show great performance in assignments.

Homework Assignment: There will be 5-6 homework assignments given to individual students. Most of them ask data management; model selection and specification; model estimation; interpretation of results; and professional presentation of results. Students have to submit their answers with Stata do files and their output attached; otherwise, their answers are considered unscientific due to lack of replicability.

Exams: There are midterm and final exam. The midterm exam consists of essay questions of reading and interpreting the Stata output. The final exam is in fact a term project where students are asked to organize their own panel data (for their theses), run a panel data model,

report and interpret its result, and then present the result professionally. Each student has to submit his/her term project proposal for data analysis describing a research question, (tentative) model specification, and panel data to be used by January 20th. A detailed guideline of the term project will be announced later.

Late Penalties: All assignments should be handed in to the instructor at the start of the class on the due date, unless instructed specifically. Late assignments will be accepted with penalty of 20% if submitted within a day after the time due. Under no circumstances will late assignments be accepted after one day of the due date without prior authorization from the instructor.

Format and Styles: All homework assignments and project papers should be written in electronic forms (Microsoft Word, OpenOffice Writer, …). Use the default format and styles (A4, default margin, Time and Roman font, 12 point, single-spaced, etc.). If you are not sure, please download the template file from the course Web page.

Organize contents clearly and logically; hit the highlights and avoid redundancy; and use proper public management jargons and grammatically correct English.

GRADING

The final grade is based on midterm exam, final exam (term project), homework assignments, and class attendance and participation. Notice that homework assignments account for the half of the final grade. Their weights are:

- Midterm exam 15%
- Final exam (term project) 20%
- Homework assignment 50%
- Class attendance 15%
- Extra credit for outstanding participation and performance (up to 10%)

Grading for this course is as follows:

96 - 100: A (4.0)	66 - 69: B- (2.5)
90 - 95: A- (3.75)	60 - 65: C (2.0)
80 - 89: B+ (3.5)	< 60 : F
70 - 79: B (3.0)	Incomplete (I), withdrawal (W)

* Students who missed 30 percent (or more) of class MAY NOT pass this course.

If you object to any grading decisions, you may appeal the grade to the instructor. The appeal must be given along with original papers, memos, and/or exams to the instructor no later than 24 hours after receiving the grade.

CLASS POLICY

Attendance and Lateness: Students should attend each class and be present when each class begins. Being 15 minutes late is considered absent from the class. Excused absences for special circumstances (e.g., sickness) may be arranged in advance and will not influence the attendance grade. There is no any formal seating chart.

Academic Dishonesty and Misconduct: Students should not only gain knowledge and skills, but also build their character. Particularly, public managers should equip themselves with

high and strict professional standards and ethics. All students should complete their own work and be evaluated based upon that work. They should avoid academic dishonesty and misconduct including *plagiarism*, *fabrication* (falsification), and *collaboration* (cheating). All aspects of IUJ' student code of conduct (see curriculum handbook) apply to this class. It is students' responsibility to be aware IUJ's policy on academic dishonesty and misconduct including sexual harassments.

Copying and pasting some parts of textbooks, journal articles, and/or Internet resources without citation involves both plagiarism and fabrication. Be honest with yourself and the instructor; *simply say that you know what you know and you don't know what you don't know* by clearly distinguishing your idea from others' ideas. Asking for other students' work or showing your work to other students is cheating. Since this course involves many calculation jobs, students should be pay special attention to this type of cheating. Free riding and/or irrelevant peer-evaluation in a group project are misconduct for the group and class. The penalties for violation include sanctions up to and including expulsion from the university.

All aspects of IUJ' student code of conduct (see students' curriculum handbook) apply to this class. It is students' responsibility to be aware IUJ's policy on academic dishonesty and misconduct including sexual harassments (http://www.iuj.ac.jp/web/iuj_section.cfm?item=090506). If you need clarification regarding this issue, contact the instructor or OAA (ofcgsir@iuj.ac.jp) immediately.

Course Feedback: Given diversity in their backgrounds, students are always encouraged to make comments and suggestions during this term in order to improve this course. Any form of communication (e.g., walk-in, phone, email, facebook, etc.) will do. No feedback will influence your grade negatively in any case.

Use of Electronic Devices: Before each class begins, students MUST turn off their cellular phones, CD/MP3/DMB players, and other electronic devices that may distract the instructor and their classmates. However, laptops or netbooks (mini laptops) are allowed for use in class unless they disturb others.

Computer Literacy: This class uses Stata but no prior knowledge of Stata is required. Students should be able to access computers in IUJ computer clusters and use Web browser (e.g., Firefox) to navigate Web sites and download materials. In particular, students SHOULD know how to download papers from IUJ journal portals (e.g., JSTOR and EBSCO). Students are expected to write memos in an electronic form (e.g., Microsoft Word, OpenOffice Writer, or PDF for LaTeX). If you are not feeling comfortable in this computing requirement, please talk to the instructor IMMEDIATELY.

IUJ Electronic mail: All students must use the university electronic mail to communicate with the instructor and other classmates. Students SHOULD peruse emails that the instructor sends in order for additional explanation and comments. Students may contact the instructor using google chat (text, audio, or video) or facebook.

Course Web Page: http://www.sonsoo.org/method/panel/ provides the latest course schedule, announcements, and various course materials including lecture notes. Students MUST visit this Web page time to time to check announcements and materials available.

READING MATERIALS

There are three required readings: some chapters in Greene (2011) and Kennedy (2008) and a copy of data analysis working paper written by the instructor. Students MUST have these textbooks or hardcopies of necessary chapters; Electronic books are not sufficient. Also some journal articles are required as shown in the weekly schedule below. Students should read all required readings before class.

Baltagi, Badi H. 2008. *Econometric Analysis of Panel Data*, 4th ed. Wiley, John & Sons. Chapter 1, 2, and 4.

Greene, William H. 2011. *Econometric Analysis*, 7th ed. Upper Saddle River, NJ: Prentice Hall, Chapter 4 and 11 (9 in 6th ed).

Kennedy, Peter. 2008. *A Guide to Econometrics*, 6th ed. Malden, MA: Blackwell Publishing. Chapter 2, 3, 4, 15, and 18.

Park, Hun Myoung. 2011. *Practical Guides To Panel Data Modeling: A Step-by-step Analysis Using Stata*. Unpublished Data Analysis Working Paper. Graduate School of International Relations, International University of Japan. http://www.sonsoo.org /method/panel/panel_iuj.pdf

Also this class recommends following books, which are available at the MLIC library reserve for 3 hour checkout. Greene (2011) and Kennedy (2008) are popular textbooks in econometrics. Baltagi (2008) and Wooldridge (2010) are classic panel data textbooks although they are a bit less readable for students without solid econometrics background. Baum (2006) and Cameron & Trivedi (2010) discuss econometric modeling methods using Stata. Long (2009) provides guides and tips of conducting scientific data analysis using Stata, most of which will be applied in this class. Mitchell (2008 and 2010) explains data management and visualization in Stata; these books will be helpful for Stata beginners.

Baltagi, Badi H. 2008. *Econometric Analysis of Panel Data*, 4th ed. Wiley, John & Sons. ISBN 978-0470518861.

Baum, Christopher F. 2006. *An Introduction to Modern Econometrics Using Stata*. College Station, TX: Stata Press. ISBN 978-1-59718-013-9.

Cameron, A. Colin, and Pravin K. Trivedi. 2010. *Microeconometrics Using Stata*, 2nd ed. College Station, TX: Stata Press. ISBN 978-1597180733.

Greene, William H. 2011. *Econometric Analysis*, 7th ed. Upper Saddle River, NJ: Prentice Hall. ISBN 978-0131395381.

Kennedy, Peter. 2008. *A Guide to Econometrics*, 6th ed. Malden, MA: Blackwell Publishing. ISBN 978-1405182577.

Long, J. Scott. 2009. *The Workflow of Data Analysis Using Stata*. College Station, TX: Stata Press. ISBN 978-1-59718-047-4.

Mitchell, Michael N. 2008. *A Visual Guide to Stata Graphics*, 2nd ed. College Station, TX: Stata Press. ISBN 978-1-59718-039-9.

Mitchell, Michael N. 2010. *Data Management Using Stata: A Practical Handbook*. College Station, TX: Stata Press. ISBN 978-1-59718-076-4.

Rabe-Hesketh, Sophia, and Anders Skrondal. 2008 *Multilevel and Longitudinal Modeling Using Stata*, 2nd ed. College Station, TX: Stata Press. ISBN 978-1-59718-040-5.

Stata Press. 2011. *Stata Longitudinal/Panel Data Reference Manual, Release 12*. College Station, TX: Stata Press. ISBN 978-1-59718-096-2.

Wonnacott, Thomas H., and Ronald J. Wonnacott. 1981. *Regression: A Second Course in Statistics*. Malabar, FL: Krieger Publishing. ISBN 978-0898749700.

Wooldridge, Jeffrey M. 2009. *Introductory Econometrics: A Modern Approach*. 4th ed. Cincinnati, OH: South-Western College. ISBN 978-0324581621.
Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press. ISBN 978-0262232586.

WEEKLY SCHEDULE

This schedule is tentative and subject to change. Students should check the latest schedule on the course Web site every week. The asterisk (*) below indicates the materials are optional.

1st Week: Panel Data and Stata Basics

Baltagi (2008) Chapter 1
Park (2011) Section 1 and 2; Mitchell (2010)
http://www.sonsoo.org/documents/DM1.pdf
http://www.sonsoo.org/documents/DM2.pdf

2nd Week: Ordinary Least Squares and Hypothesis Test

Greene (2011) Chapter 4; Kennedy (2008) Chapter 2, 3, 4; *Greene (2011) Chapter 5
Park (2008), "Hypothesis Testing and Statistical Power of a Test," Indiana University UITS Tutorial Series, http://www.indiana.edu/~statmath/stat/all/power/index.html
Lab: Park (2011) Section 4.

3rd Week: Least Squares Dummy Variable Model

Greene (2011) 6.2 and 11.3; Kennedy (2008) Chapter 15
Park (2011) Section 4
Wonnacott & Wonnacott (1981) 4.1 (pp. 104-115)
Suits, Daniel B. 1984. "Dummy Variables: Mechanics V. Interpretation." *Review of Economics & Statistics*, 66 (1): 177-180.

4th Week: Foundation for Panel Data Models

Greene (2011) Chapter 11; Kennedy (2008) Chapter 18
Park (2011) Section 3
*Woodridge (2009) Chapter 13 and 14 for panel data beginners

5th Week: Fixed Effect Model

Greene (2011) 11.4; Baltagi (2008) Chapter 2 (pp. 13-17)
Park (2011) Section 5
Lab: Baum (2006); Cameron & Trivedi (2010); Stata Press (2011)

6th Week: Random Effect Model 1

Greene (2011) 11.5.1-11.5.4; Baltagi (2008) Chapter 2 (pp. 17-22)
Park (2011) Section 6
Lab: Baum (2006); Cameron & Trivedi (2010); Stata Press (2011)

7th Week: Random Effect Model 2

Breusch, T. S., and A. R. Pagan. 1980. "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics." *Review of Economic Studies*, 47(1): 239-253.
Baltagi (2008) Chapter 4.2 (pp. 63-65)
*Rabe-Hesketh and Skrondal (2008)

8th Week: Comparing Fixed and Random Effect

Greene (2011) 11.5.5; Kennedy (2008) Chapter 18
Baltagi (2008) 4.1 (pp. 57-63) and 4.3 (pp. 72-81)
Park (2011) Section 7
Chow, Gregory C. 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions." *Econometrica*, 28 (3): 591–605.
Hausman, J. A. 1978. "Specification Tests in Econometrics." *Econometrica*, 46(6): 1251-1271.
*Swamy, P. A. V. B. 1970. "Efficient Inference in a Random Coefficient Regression Model." *Econometrica*, 38: 311-323.
*Wonnacott & Wonnacott (1981) Chapter 7.

9th Week: Nonspherical Disturbances and Others

Greene (2011) 11.6
*Baltagi (2008) Chapter 5

10th Week: Presenting Panel Data Models

Park (2011) Section 8; Mitchell (2008)
Students' presentations

Final exam (term project)